

Wireless Compressive Sensing for Energy Harvesting Sensor Nodes

Gang Yang, Vincent Y. F. Tan, Chin Keong Ho, See Ho Ting, and Yong Liang Guan

Abstract

We consider the scenario in which multiple sensors send spatially correlated data to a fusion center (FC) via independent Rayleigh-fading channels with additive noise. Assuming that the sensor data is sparse in some basis, we show that the recovery of this sparse signal can be formulated as a compressive sensing (CS) problem. To model the scenario in which the sensors operate with intermittently available energy that is harvested from the environment, we propose that each sensor transmits independently with some probability, and adapts the transmit power to its harvested energy. Due to the probabilistic transmissions, the elements of the equivalent sensing matrix are not Gaussian. Besides, since the sensors have different energy harvesting rates and different sensor-to-FC distances, the FC has different receive signal-to-noise ratios (SNRs) for each sensor. This is referred to as the *inhomogeneity* of SNRs. Thus, the elements of the sensing matrix are also not identically distributed. For this unconventional setting, we provide theoretical guarantees on the number of measurements for reliable and computationally efficient recovery, by showing that the sensing matrix satisfies the restricted isometry property (RIP), under reasonable conditions. We then compute an achievable system delay under an allowable mean-squared-error (MSE). Furthermore, using techniques from large deviations theory, we analyze the impact of inhomogeneity of SNRs on the so-called k -restricted eigenvalues, which governs the number of measurements required for the RIP to hold. We conclude that the number of measurements required for the RIP is not sensitive to the inhomogeneity of SNRs, when the number of sensors n is large and the sparsity of the sensor data (signal) k grows slower than the square root of n . Our analysis is corroborated by extensive numerical results.

Index Terms

Wireless compressive sensing, Energy harvesting, Restricted isometry property, Compressive sensing, Wireless sensor networks, Rayleigh-fading channels, Large deviations

G. Yang, S. H. Ting and Y. L. Guan are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore (e-mail:yang0305@e.ntu.edu.sg; {shting, eylguan}@ntu.edu.sg). G. Yang is supported in part by the Advanced Communications Research Program DSOCL06271, a research grant from the Directorate of Research and Technology (DRTech), Ministry of Defence, Singapore.

V. Y. F. Tan and C. K. Ho are with the Institute for Infocomm Research, A*STAR, Singapore (e-mail: {tanyfv, hock}@i2r.a-star.edu.sg). V. Y. F. Tan is also with the Department of Electrical and Computer Engineering, National University of Singapore.

I. INTRODUCTION

The lifetimes of conventional wireless sensor networks (WSNs) are limited by the total energy available in the batteries. It is inconvenient to replace batteries periodically, or even impossible when sensors are deployed in harsh conditions, e.g., in toxic environments or inside human bodies. Energy harvesting of ambient energy such as solar, wind, thermal and piezoelectric energy, appears as a promising alternative to a fixed-energy battery, to prolong the lifetime and offer potentially maintenance-free operation for WSNs [1], [2]. Compared to limited but reliable power supply from conventional batteries, energy harvesters provide a virtually perpetual but unreliable energy source. Moreover, the sensors typically have different *energy harvesting rates*, due to varying harvesting conditions such as the spread of sunlight and difference in wind speeds.

This paper addresses the problem of data transmission in energy harvesting WSNs (EH-WSNs). We assume that energy harvesting sensors are deployed to monitor some physical phenomenon in space, e.g., temperature, toxicity of gas. Data collected from sensors are sent to the fusion center (FC). The data are typically correlated, and well approximated by a sparse vector in an appropriate transform (e.g., the Fourier transform). Recent developments in compressive sensing (CS) theory provide efficient methods to recover sparse signals from limited measurements [3]. CS theory states that if the sensing matrix satisfies the restricted isometry property (RIP), a small number of measurements (relative to the length of the data vector) is sufficient to accurately recover the sparse data. This advantage of CS potentially allows us to reduce the total number of transmissions, and this is particularly important for data transmission in bandwidth-limited wireless channels.

The accurate estimation of the sensor data by the FC has recently been addressed by using CS techniques in the literature. In [4], Haupt *et. al* presented a sensing scheme based on phase-coherent transmissions for all sensors. However, [4] made two practically limiting assumptions. First, it assumed that there was no channel fading, and path losses for all sensors were identical. Second, the transmissions from all sensors were synchronized such that signals arrived in phase at the FC. In [5], Aeron *et. al* derived information theoretic bounds on sensing capacity of sensor networks under a fixed signal-to-noise ratio (SNR) for all sensors. In contrast, [6] proposed a sparse approximation method in non-fading channels, which adapted a sensor's sensing activity according to its energy availability. In [7], Xue *et. al* successively applied CS in the spatial domain and the time domain, under a fixed SNR for all sensors. In [8], Fazel *et. al* proposed a random access scheme in underwater sensor networks. Each activated sensor picked a uniformly-distributed delay

to transmit. By simply discarding the colliding data packets from concurrent medium access, the FC used a CS decoder to recover the sensor data based only on the successfully received packets. Thus, the scheme did not exploit packet collisions for data recovery.

Since sensors are placed at different locations, it is commonly assumed that the sensors transmit data over independent but nonidentical channels with different fading conditions. Different energy harvesting rates also lead to different transmit powers and hence different (receive) SNRs. We refer to this generally as the *inhomogeneity* of SNRs. The application of *wireless compressive sensing* to the scenario of inhomogeneous SNRs has, to the best of our knowledge, not been studied in the literature. We define the *system delay* as the number of concurrent sensor-to-FC transmissions (or channel uses) for estimating one data vector (among sensors). We aim to reduce the system delay, while ensuring a target estimation accuracy. Surprisingly, we observe that the required number of measurements for accurate recovery m is not overly sensitive to the inhomogeneity of SNRs provided that the number of sensors n is large and the sparsity of the data vector k grows slower than \sqrt{n} . This motivates us to further investigate the impact of inhomogeneity of SNRs, based on the recovery performance in terms of RIP.

The three main contributions are summarized as follows.

- 1) We first present an efficient transmission scheme, which features probabilistic transmission by sensor nodes. In each time slot, every sensor locally decides to transmit with some probability, and adjusts the transmit power according to its energy availability. The FC thus receives a linear combination of signals that are transmitted from a random subset of sensors. The transmissions over successive time slots result in a sensing matrix which is effectively achieved through the mixing of signals in wireless channels.
- 2) Second, we prove that the FC can recover the data accurately, if the total number of transmissions (or measurements) m exceeds

$$O\left(k \frac{\rho_{\max}(k)}{\rho_{\min}(k)} \log \frac{n}{k}\right), \quad (1)$$

where n is the number of sensors, k is the sparsity of the sensor data, and $\rho_{\max}(k)$ and $\rho_{\min}(k)$ are respectively the maximum and minimum k -restricted eigenvalues (see definition in (15)) of a Gram matrix which depend on the inhomogeneity of SNRs. Different from previous works on CS, our bound depends explicitly on the ratio $\rho_{\max}(k)/\rho_{\min}(k)$, which is the k -restricted condition number of the Gram matrix. Based on this result, we also compute the achievable

system delay subject to a desired recovery accuracy.

- 3) Third, we analyze the impact of inhomogeneity of SNRs on the required number of measurements, in terms of $\rho_{\max}(k)$ and $\rho_{\min}(k)$. We model the signal powers of the sensors as independent truncated Gaussians. By using the theory of large deviations, we show that both $\rho_{\max}(k)$ and $\rho_{\min}(k)$ concentrate around one (for all constant k) in large n regime, and the rate of convergence to one depends on the inhomogeneity of SNRs. This allows us to explain the observation that the inhomogeneity of SNRs does not significantly affect the number of measurements required for the RIP to hold.

This remainder of this paper is organized as follows: Section II provides a description of the system model. Section III presents a new wireless compressive sensing scheme. Section IV details the main results on the RIP, the achievable system delay and investigates the impact of inhomogeneity of SNRs. Section V provides the simulation results. Section VI concludes this paper. The proofs for the RIP result and the result on the impact of inhomogeneity of SNRs are given in Section VII.

We adopt the following set of notation in this paper: lower case letters denotes deterministic scalars, and lower case Greek letters for constants or angles. Boldface upper case and boldface lower case refer to matrices and (column) vectors, respectively. We use upper case letters to denote random variables. Sets are denoted with calligraphic font (e.g., \mathcal{V}). The cardinality of a finite set \mathcal{V} is denoted as $|\mathcal{V}|$. The n -order identity matrix is denoted by \mathbf{I}_n . We also use \mathbb{R}^n and \mathbb{C}^n to denote the n -dimensional real and complex Euclidean spaces respectively.

II. SYSTEM MODEL

Consider a wireless sensor network that consists of n energy harvesting sensor nodes and a FC. Sensors transmit their data to the FC via a shared multiple-access channel (MAC). We consider slotted transmissions by first considering a single snapshot of the spatial-temporal field. Assuming the sensor data \mathbf{s} is compressible, we can model it as being sparse with respect to (w.r.t.) to a fixed orthonormal basis $\{\psi_j \in \mathbb{C}^n : j = 1, \dots, n\}$, i.e.,

$$\mathbf{s} = \Psi \mathbf{x} = \sum_{j=1}^n \psi_j x_j, \quad (2)$$

where $\mathbf{x} \in \mathbb{C}^n$ has at most $k < \lfloor n/2 \rfloor$ non-zero components and $\lfloor \cdot \rfloor$ is the floor operation.

We assume a flat-fading channel with complex-valued channel coefficients h_{ij} , where $1 \leq i \leq m$ denotes the slot index and $1 \leq j \leq n$ denotes the sensor index. The channel remains constant in each

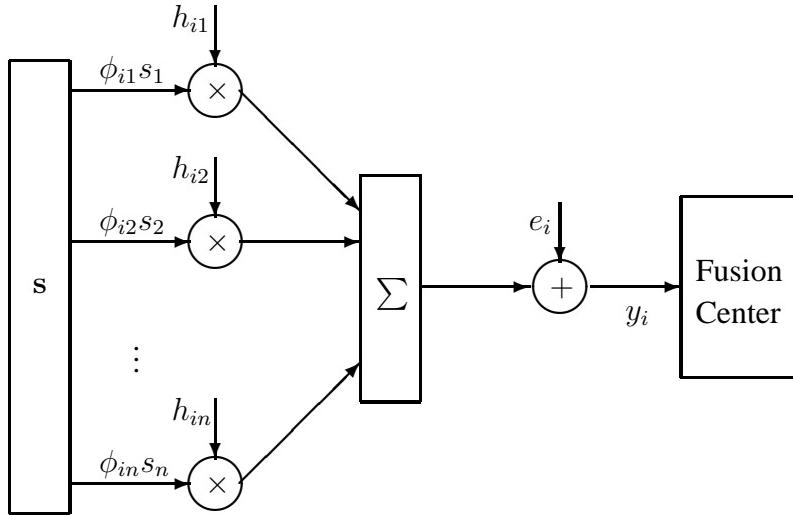


Fig. 1. The MAC communication structure for WSNs in the i -th time slot. The signals that are concurrently transmitted from sensors to the FC are linearly combined over the air.

slot. We further assume a Rayleigh-fading channel, hence the channel coefficients for different slots are independent and identically distributed (i.i.d.) according to the complex Gaussian distribution.

We propose that sensors concurrently transmit to the FC in a probabilistic manner, such that the signals from sensors are linearly combined over the air. Sensor j multiplies its datum s_j by some random amplitude ϕ_{ij} (to be defined in (4)), then transmits in the i -th time slot. The FC thus receives

$$y_i = \sum_{j=1}^n h_{ij} \phi_{ij} s_j + e_i,$$

where e_i is a noise term (not necessarily Gaussian). After m time slots, the FC receives the measurement vector

$$\mathbf{y} = (\mathbf{H} \odot \boldsymbol{\Phi})\mathbf{s} + \mathbf{e} = \mathbf{Z}\mathbf{s} + \mathbf{e} = \mathbf{Z}\Psi\mathbf{x} + \mathbf{e}, \quad (3)$$

where the matrix $\mathbf{Z} = \mathbf{H} \odot \boldsymbol{\Phi}$, and the operation \odot is the element-wise product of two matrices. We assume all noise components are independent, zero mean and have variance σ^2 . The signal model over one slot is illustrated in Fig. 1.

From the perspective of signal recovery, we want to estimate \mathbf{x} or equivalently \mathbf{s} , from \mathbf{y} , such that the *mean-squared-error* (MSE) $\mathbb{E}\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2$ does not exceed some threshold ϵ . Also, we would like to estimate the sparse vector using minimum network resources (i.e., channel uses), due to limited channel resources. Thus, given a fixed number of sensors n and an ϵ , our objective is to design a transmission scheme that minimizes the number of sensor-to-FC transmissions m .

Different from [6], [9], we consider Rayleigh-fading channels, and adopt concurrent transmissions in a probabilistic manner. Moreover, the SNRs of different sensors are considered to be different, compared to the fixed SNR case in the literature [5], [7], [9].

III. ENERGY-AWARE WIRELESS COMPRESSIVE SENSING

In Section III-A, we first provide a CS perspective for the signal model in (3). Then in Section III-B, we present an energy-aware wireless compressive sensing scheme. By taking into account the inhomogeneity of SNRs, we also derive the probability distribution function (pdf) of elements in the random matrix \mathbf{Z} in Section III-C, which will be used to show the RIP in Section IV-A.

A. A Compressive Sensing Perspective

Since we assume the data vector \mathbf{x} is sparse in some basis, it seems natural to adopt a CS method to recover \mathbf{x} . The over-the-air combination via the channel matrix \mathbf{H} contributes to the effective equivalent sensing matrix \mathbf{Z} in (3). However, there are two differences from the conventional CS setup that make the analysis more challenging.

- Due to probabilistic transmissions, the elements of the sensing matrix \mathbf{Z} are not Gaussian.
- Since sensors have different energy harvesting rates and different sensor-to-FC distances, the FC has different receive SNRs for all sensors. Thus, the elements of the sensing matrix \mathbf{Z} are also not identically distributed.

The proposed transmission scheme calls for the analysis of non-Gaussian non-i.i.d. sensing matrices. Hence, we need to analyze the system performance in a more intricate way that differs from conventional CS problems. The key technique we employ is to show that the elements of the sensing matrix \mathbf{Z} are sub-Gaussian, and make use of new results on sub-Gaussian random matrices.

B. Energy-Aware Wireless Compressive Sensing

We consider only the energy consumption for wireless transmissions, by assuming the energy consumption on sensing is negligible. The energy harvesting rate varies over sensors. For simplicity, we assume that each sensor allocates the same power for all slots. Let E_j be the accumulated harvested energy that is available for sensor j to transmit in each slot. We perform energy-aware wireless transmissions taking into account the different available energy. It is noted that a *causal*

energy constraint that comes from energy harvesting should be satisfied, i.e., energy that is consumed for transmissions can not exceed the energy available in each slot.

Set a probability $p \in (0, 1]$ and a squared-amplitude $b_j > 0$. Let Φ in (3) be a *selection-and-weight* (SW) matrix, whose elements are independently generated according to the random variable

$$\phi_{ij} = \begin{cases} +\sqrt{b_j} & \text{w.p. } p/2 \\ 0 & \text{w.p. } 1-p, \\ -\sqrt{b_j} & \text{w.p. } p/2 \end{cases} \quad \forall i = 1, 2, \dots, m. \quad (4)$$

That is, the sensor j transmits with probability p with an amplitude of $\sqrt{b_j}$, and the actual value is positive or negative with equal probability. Given available energy E_j , we choose b_j such that¹

$$pb_j \leq E_j, \quad \forall j = 1, 2, \dots, n. \quad (5)$$

Clearly, each entry ϕ_{ij} is zero mean and has variance pb_j . The causal energy constraint is satisfied in expectation, i.e., $\mathbb{E}(\phi_{ij}^2) = pb_j \leq E_j$. This allows us to save energy to be used for future transmissions. The energy-saving feature can be crucial in the scenario where the energy harvesting rates are fluctuating over several snapshots of the spatial-temporal field. It is, however, beyond the scope of this paper to optimize for the b_j 's.

In [6], all sensors consume the same amount of energy for transmissions. In contrast, each sensor here adapts the transmit power to its available energy via the above-designed SW matrix. Furthermore, the SW matrix randomly selects the sensors to transmit, and weighs the data according to the sensors' harvested energy. In each time slot, a subset of sensors are selected at random to perform transmissions and over-the-air combination. The selections are performed in a distributed manner at each sensor node, since each node separately decides the slots that it transmits in. We couple random sensor selection and energy-aware transmission by the choice of the SW matrix.

Recall the signal model in (3), i.e., $\mathbf{y} = \mathbf{Z}\Psi\mathbf{x} + \mathbf{e}$. With the knowledge² of the matrix \mathbf{Z} and the sparsity-inducing basis Ψ , the FC can implement CS decoding to recover sparse coefficients $\hat{\mathbf{x}}$ and obtain the estimated data vector $\hat{\mathbf{s}} = \Psi\hat{\mathbf{x}}$.

¹The quantity b_j can be written more generally as $b_{i,j}$, which means the transmit powers for different slots are different. To reduce the complexity of processing, we allocate the same power to all the slots.

²The assumption that the FC knows \mathbf{Z} and Ψ is reasonable, because the FC can perform channel estimation from preambles, and obtain the information on the amount of harvested energy via feedback. The channel and energy information is used for generating SW matrix from a predefined set of SW matrices. The global parameters like m and p can be broadcasted to all sensors.

C. Probability Distribution Analysis and Equivalent Normalized Signal Model

Consider the signal model in (3). Denote each element in \mathbf{Z} as $Z_{ij} = h_{ij}\phi_{ij} = Z_{ij}^R + jZ_{ij}^I$, where $Z_{ij}^R \triangleq h_{ij}^R\phi_{ij}$, and $Z_{ij}^I \triangleq h_{ij}^I\phi_{ij}$. Note that elements of the matrix \mathbf{H} are assumed to be independent, and each element h_{ij} has independent real and imaginary components. Also the matrix Φ consists of independent elements. All elements of matrix \mathbf{Z} are thus independent, and have independent real and imaginary components. As such, it suffices to analyze the probability distribution of the real component, since the analysis is similar for the imaginary component. The marginal pdf of Z_{ij}^R can be shown to be

$$f_{Z_{ij}^R}(z) = \frac{1}{\sqrt{b_j}} f_{H_j^R} \left(\frac{z}{\sqrt{b_j}} \right) \cdot \frac{p}{2} + \frac{1}{\sqrt{b_j}} f_{H_j^R} \left(-\frac{z}{\sqrt{b_j}} \right) \cdot \frac{p}{2} + (1-p) \cdot \delta(z), \quad (6)$$

where $f_{H_j^R}(\cdot)$ is the pdf of channel coefficient of sensor j , and $\delta(\cdot)$ is the Dirac delta function. For the sake of brevity, we define a new pdf as follows.

Definition 1. A random variable X follows a mixed Gaussian distribution, denoted as $X \sim \tilde{\mathcal{N}}(\mu, \nu^2, p)$, if its pdf has the following form

$$f_X(x) = p \frac{1}{\sqrt{2\pi\nu^2}} \exp \left(-\frac{(x-\mu)^2}{2\nu^2} \right) + (1-p)\delta(x), \quad (7)$$

where $p \in (0, 1]$ is the mixing parameter. The corresponding complex mixed Gaussian distribution, assuming the real and imaginary components are independent, is denoted as $\tilde{\mathcal{N}}_c(\mu, \nu^2, p)$.

Assuming Rayleigh-fading channels, all elements in the channel matrix \mathbf{H} are independent, zero mean and follow Gaussian distributions. Note that due to different fading channels for the sensors, the matrix \mathbf{H} has column-dependent variances, where the j -th column follows a Gaussian distribution with variances ν_j^2 . From (6) and (7), the marginal pdf of Z_{ij}^R can be expressed as

$$f_{Z^R}(z) = p \frac{1}{\sqrt{\pi\nu_j^2 b_j}} \exp \left(-\frac{z^2}{\nu_j^2 b_j} \right) + (1-p)\delta(z). \quad (8)$$

Thus, we have $Z^R \sim \tilde{\mathcal{N}}(0, \nu_j^2 b_j / 2, p)$.

Recall that $\mathbf{Z} = \mathbf{H} \odot \Phi$. Let $\mathbf{H} = \tilde{\mathbf{H}}\Gamma_H$ and $\Phi = \tilde{\Phi}\Gamma_\Phi$, where $\Gamma_H = \text{diag}\{\nu_1, \nu_2, \dots, \nu_n\}$ and $\Gamma_\Phi = \text{diag}\{\sqrt{pb_1}, \sqrt{pb_2}, \dots, \sqrt{pb_n}\}$. Then we can decompose the matrix \mathbf{Z} as follows

$$\mathbf{Z} = \sqrt{m} \tilde{\mathbf{Z}} \Gamma, \quad (9)$$

where we denote $\tilde{\mathbf{Z}} = \tilde{\mathbf{H}} \odot \tilde{\Phi}$ and $\Gamma = \Gamma_H \Gamma_\Phi$. Let $\Gamma = \text{diag}\{\sqrt{\gamma_1}, \sqrt{\gamma_2}, \dots, \sqrt{\gamma_n}\}$, where the receive signal power of sensor j is³ $\gamma_j = pb_j \nu_j^2$. We term the diagonal elements of Γ a signal power pattern. The γ_j 's are generally all different (i.e., inhomogeneous signal powers), and this directly leads to the inhomogeneous (receive) SNRs. We note that all elements of the matrix $\tilde{\mathbf{Z}}$ are i.i.d. mixed Gaussian random variables, i.e., $\tilde{Z} \sim \tilde{\mathcal{N}}_c(0, 1/(pm), p)$ and $\tilde{Z}^R \sim \tilde{\mathcal{N}}(0, 1/(2pm), p)$.

Using the equivalent expression in (9), we rewrite the signal model in (3) as

$$\mathbf{y} = \sqrt{m} \tilde{\mathbf{Z}} \Gamma \Psi \mathbf{x} + \mathbf{e}, \quad (10)$$

where the matrix Ψ is a unitary matrix. The distinct signal powers in Γ are spread along sparsity-inducing basis vectors (i.e., columns of Ψ).

A matrix (or more correctly, a *sequence* of matrices) is said to be *standard column regular* if all elements are uniformly bounded by some constant [10]. For analytical convenience, we normalize the matrix $\Gamma \Psi$ to be standard column regular. The normalization constant is $\|\Gamma \Psi\|_F / \sqrt{n} = \sqrt{P_{\text{ave}}}$, where $P_{\text{ave}} = \sum_{j=1}^n pb_j \nu_j^2 / n$ denotes the average (receive) signal power in one time slot. Then the normalized matrix

$$\Sigma = \Gamma \Psi / \sqrt{P_{\text{ave}}} \quad (11)$$

has bounded spectral norm. By dividing both sides of (10) by $\sqrt{m P_{\text{ave}}}$, we obtain the normalized signal model

$$\tilde{\mathbf{y}} = \tilde{\mathbf{Z}} \Sigma \mathbf{x} + \tilde{\mathbf{e}} = \mathbf{A} \mathbf{x} + \tilde{\mathbf{e}}, \quad (12)$$

where all noise components are independent, zero mean and have normalized variance $\tilde{\sigma}^2 \triangleq \sigma^2 / (m P_{\text{ave}})$. The average SNR is defined as

$$\text{SNR}_{\text{ave}} \triangleq \frac{P_{\text{ave}}}{\sigma^2} = \frac{p}{n \sigma^2} \sum_{j=1}^n b_j \nu_j^2. \quad (13)$$

IV. MAIN RESULTS

Having derived the probability distribution of elements of the matrix \mathbf{Z} in Section III-C, we recall the definition of RIP [11] and state our main result, that is Theorem 1, in Section IV-A. The engineering implication of Theorem 1, and in particular the tradeoff between the achievable

³The receive signal power depends on both the channel condition (i.e., the variance of fading coefficients ν_j^2 and the average transmit power pb_j) that is governed by the accumulated harvested energy.

system delay and the allowable MSE, will be discussed in IV-B. Finally we analyze the effect of inhomogeneity of SNRs on RIP and the required number of measurements in Section IV-C.

A. Restricted Isometry Property

It is well established in CS theory that a sufficient condition for accurate and efficient recovery (via convex optimization) is that the sensing matrix satisfies the RIP. A matrix \mathbf{A} is said to satisfy RIP of order k , if there exists a $\delta_k \in (0, 1)$ such that

$$(1 - \delta_k) \|\mathbf{x}\|_2^2 \leq \|\mathbf{Ax}\|_2^2 \leq (1 + \delta_k) \|\mathbf{x}\|_2^2 \quad (14)$$

holds for all k -sparse vectors \mathbf{x} . The smallest constant δ_k satisfying (14) is known as the *restricted isometry constant* (RIC) [11]. When the sensing matrix \mathbf{A} is random, the inequality should hold with overwhelming probability that approaches one as n grows. Many families of random matrices, e.g., i.i.d. Gaussian random matrices and Bernoulli random matrices are known to satisfy the RIP [11], [12]. As a result, to evaluate the recovery performance, all we have to show is that the sensing matrix \mathbf{A} in our scheme also obeys RIP with overwhelming probability.

The RIP requires that the sensing matrix \mathbf{A} preserves the Euclidean norm of sparse vectors well. For the signal model in (12), the entries in $\tilde{\mathbf{Z}}$ are i.i.d. *sub-Gaussian* random variables (See Definition 6 in Section VII-A). It is known that random matrices (with sufficiently many rows and) with i.i.d. sub-Gaussian entries approximately preserve the Euclidean norm of sparse vectors with high probability [13]. Since $\mathbf{A} = \tilde{\mathbf{Z}}\Sigma$, we need to analyze the norm-preserving property of Σ . To do so, we define the *k -restricted extreme eigenvalues* of the Gram matrix $\Sigma^*\Sigma$ as

$$\begin{aligned} \rho_{\max}(k) &= \max_{\mathbf{v}: \|\mathbf{v}\|_0 \leq k, \|\mathbf{v}\|_2=1} \|\Sigma\mathbf{v}\|_2^2, \\ \rho_{\min}(k) &= \min_{\mathbf{v}: \|\mathbf{v}\|_0 \leq k, \|\mathbf{v}\|_2=1} \|\Sigma\mathbf{v}\|_2^2, \end{aligned} \quad (15)$$

where $\mathbf{v} \in \mathbb{C}^n$, and the “ l_0 -norm” $\|\mathbf{v}\|_0$ refers to the number of non-zero elements of \mathbf{v} . The extreme eigenvalues will be used to understand how the inhomogeneous SNRs affects the RIP.

Lemma 1. The following bounds on $\rho_{\max}(k)$ and $\rho_{\min}(k)$ hold:

$$1 \leq \rho_{\max}(k) \leq k, \quad 0 \leq \rho_{\min}(k) \leq 1. \quad (16)$$

Proof: Fix a vector $\mathbf{v} \in \mathbb{C}^n$ such that $\|\mathbf{v}\|_2 = 1$ and $\|\mathbf{v}\|_0 = k$. Let $\mathcal{T} \subset \{1, \dots, n\}$ with

$|\mathcal{T}| \leq k$ be the support of \mathbf{v} . Let $\Sigma_{\mathcal{T}} \in \mathbb{C}^{n \times |\mathcal{T}|}$ be the submatrix of Σ with column indices \mathcal{T} . Denote the eigenvalues of the Gram matrix $\Sigma_{\mathcal{T}}^* \Sigma_{\mathcal{T}}$ by $\lambda_1 \geq \dots \geq \lambda_k \geq 0$. Due to the normalization in (11), the trace of $\Sigma_{\mathcal{T}}^* \Sigma_{\mathcal{T}}$ is $\sum_{j=1}^k \lambda_j = k$. This implies that the largest eigenvalue is at least one and at most k . Similarly, the smallest eigenvalue is no larger than one. \blacksquare

We note that the sparsity level k is usually much smaller than the number of sensors n in large-scale WSNs. We further assume $\rho_{\max}(k) \in [1, 2]$ in the following. This simplifies some of the mathematical arguments. We analytically and numerically verify this claim in Section IV-C. To state our main theoretical result cleanly, we define two quantities that depend on Σ and k as follows

$$\begin{aligned}\xi_k(\Sigma) &\triangleq \max \{1 - \rho_{\min}(k), \rho_{\max}(k) - 1\}, \\ \zeta_k(\Sigma) &\triangleq \max \left\{0, \frac{2 - \rho_{\max}(k) - \rho_{\min}(k)}{\rho_{\max}(k) - \rho_{\min}(k)}\right\}.\end{aligned}\tag{17}$$

Since $\rho_{\max}(k) \in [1, 2]$, we have⁴ $\xi_k, \zeta_k \in [0, 1]$. Let $\vartheta_k = (1 + \zeta_k)\rho_{\max}(k) - 1$. Given $\delta_k \in (\xi_k, 1)$, for convenience, we map δ_k to a “modified RIC” via a piecewise linear mapping as follows

$$\beta_k(\delta_k, \Sigma) \triangleq \begin{cases} 1 - (1 - \delta_k)/\rho_{\min}(k), & \delta_k \in (\xi_k, \vartheta_k) \\ (1 + \delta_k)/\rho_{\max}(k) - 1, & \delta_k \in (\vartheta_k, 1). \end{cases}\tag{18}$$

Let $\varsigma_k = 2/\rho_{\max}(k) - 1$. The inverse of $\beta_k(\delta_k, \Sigma)$ is denoted as

$$\delta_k(\beta_k, \Sigma) \triangleq \begin{cases} 1 - (1 - \beta_k)\rho_{\min}(k), & \beta_k \in (0, \zeta_k) \\ (1 + \beta_k)\rho_{\max}(k) - 1, & \beta_k \in (\zeta_k, \varsigma_k). \end{cases}\tag{19}$$

In the sequel, we assume that the quantity $\xi_k(\Sigma)$ is a small positive number and it measures the inhomogeneity of the eigenvalues of $\Sigma_{\mathcal{T}}^* \Sigma_{\mathcal{T}}$ for $|\mathcal{T}| \leq k$. This implies ζ_k is small, and the deviation between β_k and δ_k is also small. The validity of this assumption will be shown both analytically and numerically in Section IV-C.

Recall that the sensing matrix $\mathbf{A} = \tilde{\mathbf{Z}}\Sigma$ in (12), where all elements of the $m \times n$ matrix $\tilde{\mathbf{Z}}$ are i.i.d. mixed Gaussian random variables, Σ is defined in (11), and n is the number of sensors. We now state our main theoretical result.

⁴The arguments of some quantities are sometimes omitted for notational convenience.

Theorem 1. Let $c_1, c_2 > 0$ be some universal constants. Given a sparsity level $k < \lfloor n/2 \rfloor$, a transmit probability $p \in (0, 1]$ and a number $\delta_k \in (\xi_k, 1)$, if the number of measurements satisfies

$$m > \frac{c_1 k \rho_{\max}(k)}{p^2 \beta_k^2 \rho_{\min}(k)} \log \frac{5en}{k}, \quad (20)$$

where $\beta_k = \beta_k(\delta_k, \Sigma)$ is defined in (18), then for any vector \mathbf{x} with support of cardinality of at most k , we have that the RIP in (14) holds with probability at least

$$1 - \exp(-c_2 mp^2 \beta_k^2 / 4). \quad (21)$$

Proof: See Section VII-A. ■

Remark 1 (Specialization to the homogeneous case). Clearly, the lower bound on the required number of measurements is $O(\frac{k \rho_{\max}(k)}{\beta_k^2 \rho_{\min}(k)} \log \frac{n}{k})$. For the homogeneous signal power pattern (i.e., the matrix Γ is a multiple of the identity matrix \mathbf{I}_n), we have $\rho_{\max}(k) = \rho_{\min}(k) = 1$ and $\beta_k = \delta_k$. Thus the lower bound reduces to $O(\frac{k}{\delta_k^2} \log \frac{n}{k})$, which coincides with the known results for i.i.d. random sensing matrices. See Theorem 5.2 in [12] and Section 1.4.4 in [13].

Remark 2 (Contribution to the RIP analysis). Due to the inhomogeneous signal power pattern, the rows \mathbf{a}_i of the sub-Gaussian sensing matrix \mathbf{A} are *non-isotropic*. To the best of our knowledge, little is known about the RIP of non-isotropic sub-Gaussian random matrices. The only relevant result is in Remark 5.40 in [13] which gives a concentration inequality of non-isotropic random sensing matrices in terms of the upper bound on the spectral norm. However, the authors did not demonstrate how the inhomogeneity affects the RIP, nor did they investigate the number of measurements required to satisfy the RIP. Theorem 1 fills this gap.

Remark 3. Theorem 1 is proved in Section VII-A by leveraging Theorem 2.1 of [14], which states that a sufficient condition for the approximate preservation of the Euclidean norm upon random linear mapping is that the number of measurements is proportional to the fourth power of the sub-Gaussian norm. In our scenario, as shown in Lemma 6, the sub-Gaussian norm bounded above by $1/\sqrt{p}$. In addition, Lemma 6 shows (using the Chernoff-bound) that the sub-Gaussian tail probability is bounded above by $pe^{-pt^2/2}$. Note that the sub-Gaussian norm is the smallest constant $\varrho > 0$ for which the sub-Gaussian tail probability is $2e^{-t^2/(2\varrho^2)}$ (Definition 6). In view of the fact that the pre-factor in our bound is p (and not 2), there is some degradation with respect to p in Theorem 1. For larger p , the degradation is reduced.

B. Achievable System Delay

The performance of wireless compressive sensing scheme is characterized by two quantities, i.e., the MSE and the system delay. The MSE performance under bounded noise is studied in the CS literature [3], [15], [16]. Note that there is often a trade-off between the two quantities. Under an allowable MSE $\epsilon > 0$, we thus analyze the *achievable system delay* $D(\epsilon)$, which is defined as

$$D(\epsilon) \triangleq \min_m m \quad \text{subject to} \quad \mathbb{E}\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \leq \epsilon. \quad (22)$$

Corollary 1. Let $p, m, n, k, \Sigma, \xi_k, \vartheta_k$ be as in Theorem 1. Let $\epsilon_{\text{th}} \triangleq 1/(0.0942 \times \text{SNR}_{\text{ave}})$. Given an allowable MSE $\epsilon > \epsilon_{\text{th}}$, with overwhelming probability (exceeding (21)), the achievable system delay is

$$D(\epsilon) = \frac{c_1 k \rho_{\max}(k)}{p^2 (\tilde{\beta}_k)^2 \rho_{\min}(k)} \log \frac{5en}{k}, \quad (23)$$

where

$$\tilde{\beta}_k(\Sigma, \epsilon) \triangleq \begin{cases} 1 - \frac{0.693 + 1/\sqrt{\epsilon \text{SNR}_{\text{ave}}}}{\rho_{\min}(k)}, & \delta_k \in (\xi_k, \vartheta_k), \\ \frac{1.307 - 1/\sqrt{\epsilon \text{SNR}_{\text{ave}}}}{\rho_{\max}(k)} - 1, & \delta_k \in (\vartheta_k, 1). \end{cases} \quad (24)$$

Proof: We start the proof by leveraging on the following lemma.

Lemma 2 (Theorem 3.2 of [15]). Let $\tilde{\mathbf{y}} = \mathbf{A}\mathbf{x} + \tilde{\mathbf{e}}$, where \mathbf{x} is a k -sparse vector in \mathbb{C}^n , $\tilde{\mathbf{e}} \in \mathbb{C}^m$ is a zero mean, white random vector whose entries have variance σ^2 . If the \mathbf{A} satisfies the RIP with RIC $\delta_k < 0.307$, then the solution $\hat{\mathbf{x}}$ to the ℓ_1 -minimization problem in CS decoder [3], [13] satisfies

$$\mathbb{E}\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \leq \frac{\sigma^2}{P_{\text{ave}}(0.307 - \delta_k)^2}. \quad (25)$$

Recall the definition of SNR_{ave} in (13). From Lemma 2, to achieve a MSE ϵ , it suffices to ensure the RIC satisfies $\delta_k^* = 0.307 - 1/\sqrt{\epsilon \text{SNR}_{\text{ave}}}$. From Theorem 1, the required minimum number of measurements such that the RIP holds with overwhelming probability is

$$m_{\min} = \frac{c_1 k \rho_{\max}(k) \log \frac{5en}{k}}{p^2 (\tilde{\beta}_k^*)^2 \rho_{\min}(k)} \quad (26)$$

where $\tilde{\beta}_k$ is given in (24). The definition of the achievable system delay establishes Corollary 1. ■

Remark 4. Note that Corollary 1 applies only to the case where the MSE ϵ is greater than the threshold ϵ_{th} . If $\epsilon \leq \epsilon_{\text{th}}$, then from (25), simple algebra reveals that $\delta_k = 0$, which implies that

the sensing matrix \mathbf{A} is a perfect isometry. Since \mathbf{A} is random, and the entries are governed by a density that is absolutely continuous w.r.t. the Lebesgue measure, this occurs with probability zero, implying that the constraint in (22) is almost surely not satisfied. Thus, in this case, we define the system delay to be ∞ .

Remark 5. As either ϵ or SNR_{ave} increases, $\tilde{\beta}_k$ increases, and thus the system delay $D(\epsilon)$ decreases. More importantly, we note from Corollary 1 that the key measure for the inhomogeneity of SNRs is the ratio $r(k) \triangleq \rho_{\max}(k)/\rho_{\min}(k) \in [1, \infty)$. The system delay increases as $r(k)$ increases from one. We hence analyze the impact of inhomogeneity of SNRs on the deviation of $\rho_{\max}(k)$ and $\rho_{\min}(k)$ from unity in Section IV-C. In addition, the system delay decreases as p increases, since SNR_{ave} defined in (13) increases as p increases. Thus, there is an inherent tradeoff between system delay and energy consumption because large p implies high transmit energy. Thus, it is always advantageous to transmit with as high a probability as possible subject to the causal energy constraint.

Example 1. Let the number of sensors $n = 500$, the sparsity level $k = 5$ and the transmit probability $p = 0.8$. These parameters imply $\rho_{\max}(k) = 1.09$, $\rho_{\min}(k) = 0.88$ (See Section IV). We plot the achievable system delay $D(\epsilon)$ against the allowable MSE ϵ , for different average SNRs in Fig. 2. We observe that beyond the MSE threshold (that depends on the average SNR), the system delay $D(\epsilon)$ decreases as either ϵ or SNR_{ave} increases, which is expected.

Remark 6. We considered the scenario in which the FC collects one data vector from all sensors in one frame. As a generalization of our setup, one can seek to minimize the total number of slots for collecting multiple data vectors. By adjusting the transmit probability in each frame, one can allocate different powers for different frames, such that both the recovery accuracy and the causal energy constraint is guaranteed. Details of this possible extension are beyond the scope of this paper.

C. Effect of Inhomogeneity

This section investigates the impact of inhomogeneity of (receive) SNRs on the number of measurements needed to satisfy the RIP. Without loss of generality, we assume all sensors have the same noise power, hence, it suffices to analyze the impact of inhomogeneity of receive signal powers. We focus on the asymptotic scenario where the number of sensors n tends to infinity and, for the ease of analysis, k is kept constant. To make the dependence on n clear, we denote $\rho_{\max}(k)$ (resp. $\rho_{\min}(k)$) as $\rho_{\max}(k, n)$ (resp. $\rho_{\min}(k, n)$). It will be shown that both $\rho_{\max}(k, n)$ and

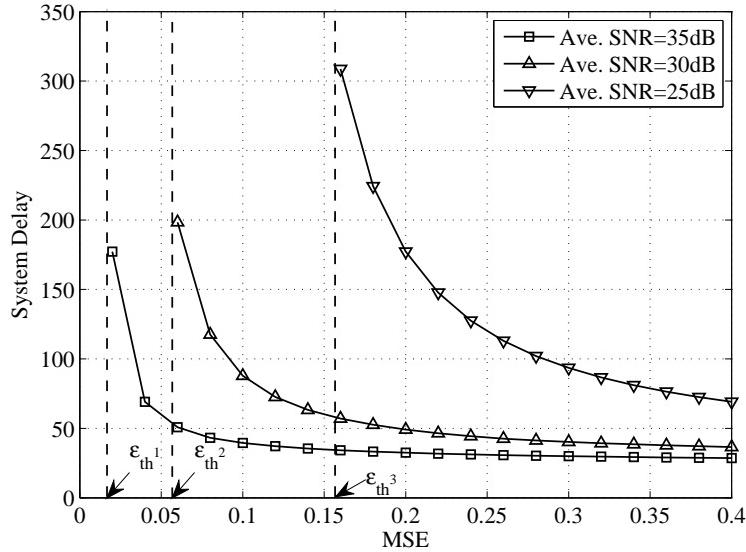


Fig. 2. Plot of achievable system delay against allowable MSE. Beyond the MSE threshold, the achievable system delay increases as either the allowable MSE or the average SNR increases.

$\rho_{\min}(k, n)$ concentrate around one when n is large, and the rate of convergence to one depends on the inhomogeneity of SNRs. This implies that the recovery performance (the required number of measurements and the probability that the RIP holds in Theorem 1) is not sensitive to the inhomogeneity of SNRs when n is large.

Let $\mathbf{w} = \Sigma \mathbf{v}$, where the unit-norm, k -sparse vector \mathbf{v} is supported on the set $\mathcal{T} \triangleq \{s_1, \dots, s_k\} \subset \{1, \dots, n\}$ and let $s_1 < \dots < s_k$. To obtain further insights, we let Ψ be the n -point discrete Fourier transform (DFT) matrix. Then the squared ℓ_2 -norm of \mathbf{w} can be expressed as follows

$$\|\mathbf{w}\|_2^2 = \frac{1}{nP_{\text{ave}}} \sum_{i=1}^n \gamma_i \left(1 + \sum_{q=1}^k \sum_{l=1, l \neq q}^k 2\text{Re} \left\{ v_{s_q} v_{s_l}^* \exp \left(\frac{-j2\pi(i-1)(s_q - s_l)}{n} \right) \right\} \right). \quad (27)$$

Since $\|\mathbf{w}\|_2^2$ is strongly influenced by the inner summation terms, we analyze the behavior of these terms more carefully in the sequel. When the signal power pattern is homogeneous, i.e., $\Gamma = \text{diag}(\sqrt{\gamma}, \dots, \sqrt{\gamma})$, we have $\|\mathbf{w}\|_2^2 = \|\Sigma \mathbf{v}\|_2^2 = 1$, hence $\rho_{\max}(k, n) = \rho_{\min}(k, n) = 1$ for all k, n .

We are interested to know how $\rho_{\max}(k, n)$ and $\rho_{\min}(k, n)$ vary with different signal powers γ_i 's. Thus, we consider a model in which the γ_i 's are i.i.d. random variables following an approximate Gaussian distribution. By varying the variance of this distribution, we are in fact varying the inhomogeneity of the signal powers. Specifically, to deal with the fact that the signal powers cannot

be negative, we use the following truncated Gaussian distribution to model the signal powers.

Definition 2. A random variable X is truncated Gaussian, denoted as $\mathcal{N}_{\text{tr}}(\mu, \omega^2)$, if its pdf is

$$g_X(x; \mu, \omega) = \frac{1}{\sqrt{2\pi}\omega(1 - Q(\mu/\omega))} \exp\left(-\frac{(x - \mu)^2}{2\omega^2}\right), \quad (28)$$

for $x \geq 0$ and 0 else, where $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$ is the Q-function of a standard Gaussian pdf.

We assume that $\gamma_i \sim \mathcal{N}_{\text{tr}}(\mu, \omega^2)$ for all $i = 1, \dots, n$ and they are mutually independent. Given μ , the “variance” ω^2 is a measure of the degree of inhomogeneity of the signal powers γ_i ’s. Also, the parameter $d \triangleq \mu/\omega$ is a measure of the homogeneity of the SNRs. If d is small (resp. large), the SNRs are less (resp. more) homogeneous. We use the exponential asymptotic notation $a_n \leq \exp(-nE)$ to mean that $\limsup_{n \rightarrow \infty} \frac{1}{n} \log a_n \leq -E$. Under the above assumptions on the statistics of the signal powers, we have the following large deviations upper bound on $\rho_{\max}(k, n)$ and $\rho_{\min}(k, n)$:

Theorem 2. Let $d \triangleq \mu/\omega$. For any $t > 0$, and any constant $1 \leq k < \lfloor n/2 \rfloor$,

$$\begin{aligned} \mathbb{P}(\rho_{\max}(k, n) > 1 + t) &\leq \exp[-nd^2 E(k, t)^2], \\ \mathbb{P}(\rho_{\min}(k, n) < 1 - t) &\leq \exp[-nd^2 E(k, t)^2], \end{aligned} \quad (29)$$

where the exponent $E(k, t)$ is defined as $E(k, t) \triangleq t/(k - 1 + \sqrt{2}t)$.

Proof: See Section VII-B. ■

Recall that Theorem 1 says that both the required number of measurements and the probability that the RIP holds depends on the ratio $r(k, n) = \rho_{\max}(k, n)/\rho_{\min}(k, n)$. From Theorem 2, we note that both $\rho_{\max}(k, n)$ and $\rho_{\min}(k, n)$ concentrate around one in the large n regime (for bounded k), and the rate of convergence to one depends on the inhomogeneity of SNRs. This allows us to conclude that for large-scale EHWSNs (relative to the signal sparsity), the inhomogeneity of SNRs does not significantly affect the RIP and the system delay, which is a surprisingly positive observation.

Remark 7. We note that $E(k, t)$ is an increasing function of t and a decreasing function of the sparsity k which is expected. Also, the exponent $d^2 E(k, t)^2$ increases with d , which means that the convergence of $\rho_{\max}(k, n)$ and $\rho_{\min}(k, n)$ to unity is faster when d is large, or equivalently, when the signal powers are more homogeneous. It is observed that $\rho_{\max}(k, n)$ is close to one in the large n regime. This validates the assumption that $\rho_{\max}(k, n) \in [1, 2]$ in Section IV-A.

Remark 8. In the preceding analysis, and particularly in Theorem 2, we assumed that k does not grow with n . Close examination of the proof shows that if $k = \lfloor n^{1/2-\lambda} \rfloor$ for any $\lambda \in (0, 1/2]$, then the probability that $\{\rho_{\max}(k, n) > 1+t\}$ still goes to zero albeit at a slower rate of $\approx \exp(-n^{2\lambda}d^2t^2)$ (not exponential in n). More precisely, we can verify that

$$\limsup_{n \rightarrow \infty} \frac{1}{n^{2\lambda}} \log \mathbb{P}(\rho_{\max}(k, n) > 1+t) \leq -d^2t^2, \quad (30)$$

and analogously for $\{\rho_{\min}(k, n) < 1-t\}$. Inequality (30) is a so-called *moderate-deviations* result [17, Sec. 3.7]. Notice that the dependencies on the homogeneity $d = \mu/\omega$ and t are similar to (29).

Remark 9. One may wonder whether Theorem 2 depends strongly on Ψ being the DFT matrix. In fact, the only property of the DFT that we exploit in the proof of Theorem 2 is its circular symmetry, i.e., each basis vector of the DFT (containing elements that are powers of the n -th root of unity) is uniformly distributed over the circle in the complex plane. Hence, certain Cesàro-sums converge to zero and the proof goes through. See (44) in Section VII-B. Thus, Theorem 2 also applies for other sparsity-inducing bases whose basis vectors have the circular symmetric property, e.g., the discrete cosine transform (DCT) or the Hadamard transform.

V. SIMULATION RESULTS

We now numerically validate our results. We set the number of sensors $n = 500$ and transmit probability $p = 0.8$. We use the truncated Gaussian distribution with $\mu = 0.2$ to model the receive signal powers, and use the basis pursuit de-noising (BPDN) algorithm [18] as the CS decoder.

First, we fix $d = 2$, which implies $\omega = \mu/d = 0.1$. Fig. 3 plots the MSE against the number of measurements (or transmissions) m for different sparsities k and different average SNRs. As expected, the MSE decreases as either k decreases or the average SNR increases. Consider the MSE level 2×10^{-3} . When the average SNR is 25 dB, the wireless compressive sensing scheme achieves a smaller system delay of $D = 68$ for $k = 5$ compared to $D = 115$ for $k = 10$. When the sparsity $k = 5$, the scheme achieves a smaller system delay of $D = 39$ for $\text{SNR}_{\text{ave}} = 30$ dB compared to $D = 68$ for $\text{SNR}_{\text{ave}} = 25$ dB.

Second, we fix $d = 2$ and the average SNR to be 25 dB. Fig. 4 compares the MSEs of the inhomogeneous SNR and the homogeneous SNR scenarios, for the sparsity levels $k = 5, 10, 20$. It is observed that in the inhomogeneous scenario, the MSE performance is slightly worse than that of the homogeneous-SNR scenario. Note that the degradation becomes larger as the sparsity k

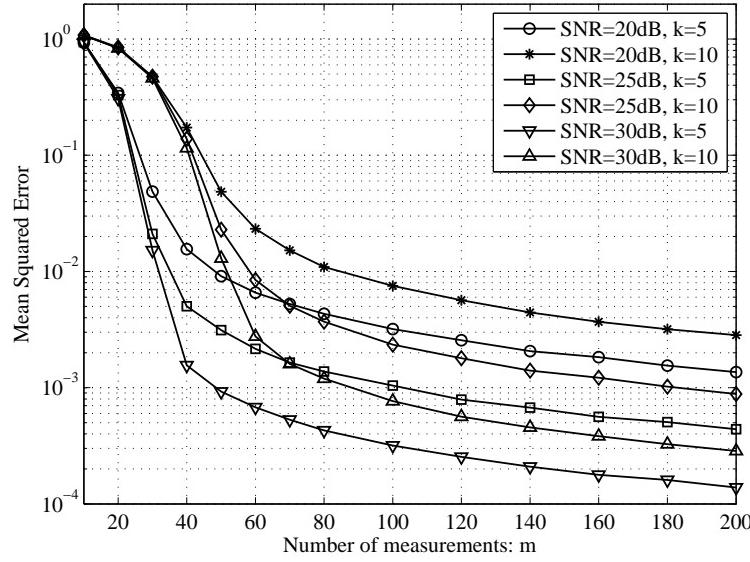


Fig. 3. Plot of MSE against the number of measurements. The MSE decreases as k decreases, or the average SNR increases.

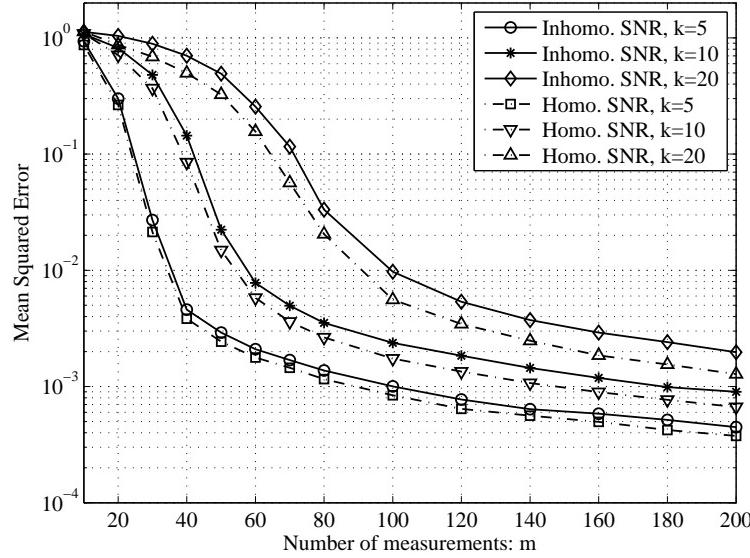


Fig. 4. Plot of MSE against the number of measurements. The MSE performance for the inhomogeneous scenario is slightly worse than that of the homogeneous-SNR scenario.

increases. This is because the convergence rate for $\rho_{\max}(k)$ and $\rho_{\min}(k)$ to one is faster if k is small relative to n . This corroborates the observation in Section IV-C.

Third, we set $d = 1, 2$ and $k = 5, 10$. Fig. 5 shows the cumulative distribution function (CDF) of

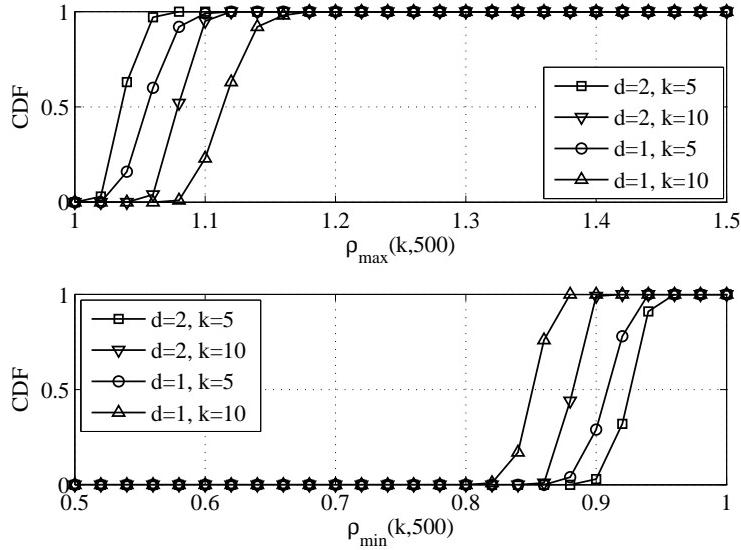


Fig. 5. CDF of $\rho_{\max}(k, 500)$ and $\rho_{\min}(k, 500)$. Both converge to one faster for more homogeneous SNRs (i.e., large d).

$\rho_{\max}(k, 500)$ and $\rho_{\min}(k, 500)$. We note that both $\rho_{\max}(k, 500)$ and $\rho_{\min}(k, 500)$ converge to one faster for larger d , or equivalently, for more homogeneous SNRs. Also, under the same inhomogeneous SNRs, both $\rho_{\max}(k, 500)$ and $\rho_{\min}(k, 500)$ converge to one faster for smaller k .

Finally, we numerically validate the asymptotic behavior of $\rho_{\max}(k, n)$ as n grows. Set $k = 5$, $d = 1, 2, 3$, respectively. Fig. 6 shows the probability that $\rho_{\max}(k, n) > 1.04$ for different n . It is observed that the logarithm of the probability decreases linearly as n grows (when n/k is large) and furthermore, the slope varies quadratically w.r.t. d , i.e., the slope is proportional to $-1, -4, -9$ for $d = 1, 2, 3$, respectively. This observation corroborates Theorem 2.

VI. CONCLUSION

In this paper, we considered the scenario in which each sensor independently decides whether or not to transmit with some probability p , and the overall transmission power (and thus p) depends on its available energy. Hence, only a subset of sensors transmits concurrently to the FC, and this exploits the spatial combination inherent in wireless channels. We use techniques from CS theory to prove a lower bound on the required number of measurements to satisfy the RIP and hence to ensure that the data recovery is both computationally efficient (and amenable to convex optimization) and accurate. We also compute an achievable system delay given an allowable MSE. Finally, we analyze the impact of inhomogeneity on the k -restricted extreme eigenvalues. These eigenvalues govern the

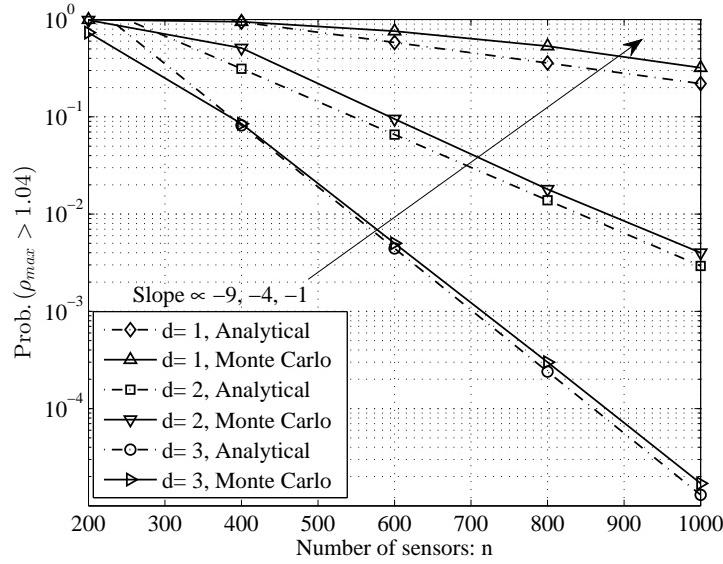


Fig. 6. Plot of the probability of $\rho_{\max}(k, n) > 1.04$ against the number of sensors. The logarithm of the probability decreases linearly as n grows, and the slope varies quadratically w.r.t. d .

number of measurements required for the RIP to hold. In large-scale EH-WSNs, we showed using large deviation techniques that the recovery accuracy and the system delay are not sensitive to the inhomogeneity of SNRs.

VII. PROOFS OF MAIN RESULTS

A. Proof of Theorem 1

Proof: Recall the signal model in (12), i.e., $\tilde{\mathbf{y}} = \tilde{\mathbf{Z}}\Sigma\mathbf{x} + \tilde{\mathbf{e}} = \mathbf{A}\mathbf{x} + \tilde{\mathbf{e}}$. The proof involves three steps. In step 1 and step 2, we prove the desired result when all quantities are real; and in step 3, we extend the result to the complex case. For the real case, we show that the matrix $\tilde{\mathbf{Z}}$ acts as isometry on the images of the sparse vector under matrix Σ , i.e., on the set $\{\Sigma\mathbf{v} : \|\mathbf{v}\|_0 \leq k, \mathbf{v} \in \mathbb{R}^n\}$. By showing the rows of $\tilde{\mathbf{Z}}$ are isotropic sub-Gaussian and by exploiting the so-called “restricted eigenvalue property” of Σ , we derive an RIP for the matrix \mathbf{A} in step 2. Before step 1, we start with the following preliminaries. Let $d(\mathbf{u}, \mathbf{v})$ be the Euclidean distance in \mathbb{R}^n .

Definition 3 (Nets, covering numbers [13]). *Consider a metric space (\mathcal{U}, d) with $\mathcal{U} \subset \mathbb{R}^n$ and a positive number ϵ . A subset $\mathcal{N}_\epsilon \subset \mathcal{U}$ is called an ϵ -net of \mathcal{U} if every point $\mathbf{u} \in \mathcal{U}$ can be approximated to within ϵ by some point $\mathbf{v} \in \mathcal{N}_\epsilon$, i.e., $d(\mathbf{u}, \mathbf{v}) \leq \epsilon$. The covering number $\mathcal{N}(\mathcal{U}, \epsilon)$ is the cardinality of the smallest ϵ -net of \mathcal{U} .*

Definition 4 (Set of sparse vectors). Let \mathcal{S}^{n-1} be the unit sphere in \mathbb{R}^n and $1 \leq k \leq n$. Define

$$\mathcal{U}_k \triangleq \{\mathbf{u} \in \mathcal{S}^{n-1} : \|\mathbf{u}\|_0 \leq k\},$$

also define the subset of the Euclidean unit ball \mathcal{B}_2^n with (at most) k -sparse vectors as

$$\tilde{\mathcal{U}}_k \triangleq \{\mathbf{u} \in \mathcal{B}_2^{n-1} : \|\mathbf{u}\|_0 \leq k\}.$$

Lemma 3 (Upper bound on covering numbers, Lemma 2.3 in [14]). Let $0 < \epsilon < 1$ and $1 \leq k \leq n$. There exists an ϵ -net of $\tilde{\mathcal{U}}_k$, namely \mathcal{N}_ϵ , whose cardinality can be upper bounded as

$$|\mathcal{N}_\epsilon| \leq \left(\frac{5}{2\epsilon}\right)^k \binom{n}{k}.$$

Definition 5 (Complexity measure [14]). The complexity of a set $\mathcal{V} \subset \mathbb{R}^n$ is defined as

$$\ell_*(\mathcal{V}) \triangleq \mathbb{E} \left[\sup_{\mathbf{v} \in \mathcal{V}} |\langle \mathbf{v}, \mathbf{u} \rangle| \right],$$

where $\langle \cdot, \cdot \rangle$ denotes inner product in \mathbb{R}^n , $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is a standard Gaussian random vector, and the supremum is over all vectors $\mathbf{v} \in \mathcal{V}$.

Given a subset $\mathcal{V} \subset \mathbb{R}^n$, we aim to measure the complexity of $\mathcal{W}(\mathcal{V})$, which is the image set of the set \mathcal{V} under a fixed linear mapping Σ . More precisely, we define

$$\mathcal{W}(\mathcal{V}) \triangleq \{\mathbf{w} \in \mathbb{R}^n : \mathbf{w} = \Sigma \mathbf{v}, \text{ for some } \mathbf{v} \in \mathcal{V}\}. \quad (31)$$

Define the complexity of $\mathcal{W}(\mathcal{V})$ as $\ell_*(\mathcal{W}(\mathcal{V})) \triangleq \mathbb{E} [\sup_{\mathbf{v} \in \mathcal{V}} |\langle \mathbf{v}, \Sigma \mathbf{u} \rangle|]$.

Lemma 4 (Upper bound on complexity measure, Lemma B.6 in [19]). Let $\mathcal{N}_{\frac{1}{2},k}$ be a $\frac{1}{2}$ -net of $\tilde{\mathcal{U}}_k$ provided by Lemma 3. Then for all $1 \leq k \leq n$, it holds that

$$\begin{aligned} \ell_* (\mathcal{W}(\mathcal{N}_{\frac{1}{2},k})) &\leq 3 \sqrt{k \rho_{\max}(k) \log \frac{5en}{k}}, \\ \ell_* (\mathcal{W}(\mathcal{U}_k)) &\leq 2 \ell_* (\mathcal{W}(\mathcal{N}_{\frac{1}{2},k})), \end{aligned} \quad (32)$$

where $\rho_{\max}(k)$ is the k -restricted maximum eigenvalue of $\Sigma^* \Sigma$ defined in (15).

Define the set

$$\mathcal{E}_k \triangleq \{\mathbf{v} \in \mathbb{R}^n : \|\Sigma \mathbf{v}\|_2 = 1, \|\mathbf{v}\|_0 = k\}, \quad (33)$$

then for $\mathcal{V} = \mathcal{E}_k$, the complexity measure of the set $\mathcal{W}(\mathcal{E}_k)$ is bounded in the following Lemma.

Lemma 5. The complexity measure of the set $\mathcal{W}(\mathcal{E}_k)$ is upper bounded as

$$\ell_*(\mathcal{W}(\mathcal{E}_k)) \leq 6 \sqrt{k \frac{\rho_{\max}(k)}{\rho_{\min}(k)} \log \frac{5en}{k}}, \quad (34)$$

where $\rho_{\max}(k)$ and $\rho_{\min}(k)$ are defined in (15).

Proof: For any vector $\mathbf{v} \in \mathcal{E}_k$ and any random vector $\mathbf{u} \in \mathbb{R}^n$, we have with probability one that

$$|\langle \mathbf{u}, \Sigma \mathbf{v} \rangle| = |\langle \mathbf{v}, \Sigma \mathbf{u} \rangle| = \|\mathbf{v}\|_2 \left| \left\langle \frac{\mathbf{v}}{\|\mathbf{v}\|_2}, \Sigma \mathbf{u} \right\rangle \right| \leq \|\mathbf{v}\|_2 \sup_{\mathbf{r} \in \mathcal{U}_k} |\langle \mathbf{r}, \Sigma \mathbf{u} \rangle|, \quad (35)$$

where the inequality follows from the definition of the set $\{\frac{\mathbf{v}}{\|\mathbf{v}\|_2} : \mathbf{v} \in \mathcal{E}_k\} \subset \mathcal{U}_k$. From Lemma 4,

$$\begin{aligned} \mathbb{E} \left[\sup_{\mathbf{v} \in \mathcal{E}_k} |\langle \mathbf{u}, \Sigma \mathbf{v} \rangle| \right] &\stackrel{(a)}{\leq} \sup_{\mathbf{v} \in \mathcal{E}_k} \|\mathbf{v}\|_2 \mathbb{E} \left[\sup_{\mathbf{r} \in \mathcal{U}_k} |\langle \mathbf{r}, \Sigma \mathbf{u} \rangle| \right] \\ &\stackrel{(b)}{\leq} 6 \sqrt{k \frac{\rho_{\max}(k)}{\rho_{\min}(k)} \log \frac{5en}{k}}. \end{aligned} \quad (36)$$

where (a) comes from (35) and (b) follows from Lemma 4 and the definitions in (15). ■

Step 1: Isometry on the images of sparse vectors. We consider the case in which the sensor data and all matrices are real. In this step, we first show that all row vectors in matrix $\tilde{\mathbf{Z}}$ are isotropic sub-Gaussian (see Definition 7 below) in Lemma 6. Then we use Lemma 5 to obtain an isometry on the images of sparse vectors.

Definition 6 (sub-Gaussian random variables [13]). *Let X be a zero mean random variable that has unit variance. It is sub-Gaussian if for any $t \geq 0$, there exist a positive number ϱ such that*

$$\mathbb{P}(|X| \geq t) \leq 2 \exp \left(-\frac{t^2}{2\varrho^2} \right).$$

The sub-Gaussian norm $\|X\|_{\psi_2}$ is the smallest number ϱ for which the above inequality holds.

Definition 7 (Isotropic sub-Gaussian random vectors [13]). *Let \mathbf{u} be a random vector in \mathbb{R}^n . If $\mathbb{E}[\mathbf{u}\mathbf{u}^T] = \mathbf{I}_n$, then \mathbf{u} is called isotropic. The random vector \mathbf{u} is sub-Gaussian with constant α if*

$$\sup_{\mathbf{r} \in \mathbb{R}^n : \|\mathbf{r}\|_2=1} \|\langle \mathbf{u}, \mathbf{r} \rangle\|_{\psi_2} < \alpha.$$

Lemma 6. Let $\mathbf{u} \in \mathbb{R}^n$ be a random vector with i.i.d. elements, each distributed as $\tilde{\mathcal{N}}(0, 1/p, p)$. Then \mathbf{u} is isotropic sub-Gaussian with constant $\alpha = c_0/\sqrt{p}$, where c_0 is an absolute constant.

Proof: Since all elements in \mathbf{u} are independent zero mean random variables, and has unit variance, we have $\mathbb{E}[\mathbf{u}\mathbf{u}^T] = \mathbf{I}_n$. Let $X \sim \tilde{\mathcal{N}}(0, 1/p, p)$ be a mixed Gaussian random variable with pdf defined in (7). Then, we have for every $t \geq 0$ that

$$\begin{aligned}\mathbb{P}(|X| > t) &= 2 \int_{\sqrt{pt}}^{\infty} p \cdot \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2}{2}\right) dx \\ &\stackrel{(a)}{\leq} pe^{-pt^2/2} \stackrel{(b)}{\leq} 2e^{-pt^2/2},\end{aligned}$$

where (a) follows from the Chernoff bound on Gaussian Q -function, and (b) from $p \in (0, 1]$. Hence, the sub-Gaussian norm of X is bounded above by $1/\sqrt{p}$. From Lemma 5.24 in [13], we have that the vector \mathbf{u} is sub-Gaussian with constant $\alpha = c_0/\sqrt{p}$, where c_0 is an absolute constant. ■

Recall that the signal model is $\tilde{\mathbf{y}} = \tilde{\mathbf{Z}}\Sigma\mathbf{x} + \tilde{\mathbf{e}}$. We note that all elements in matrix $\tilde{\mathbf{Z}}$ are i.i.d. with distribution $\tilde{\mathcal{N}}(0, 1/(mp), p)$. Then Lemma 6 implies that all row vectors of scaled matrix $\sqrt{m}\tilde{\mathbf{Z}}$ are independent, and isotropic sub-Gaussian with constant $\alpha = c_0/\sqrt{p}$. The key idea to prove Theorem 1 is to apply one result in [14], which is given without proof as follows.

Lemma 7 (Theorem 2.1 in [14]). Set $1 \leq m \leq n$ and $0 < \beta < 1$. Let \mathbf{b} be an isotropic sub-Gaussian random vector on \mathbb{R}^n with constant $\alpha \geq 1$. Let $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$ be independent copies of \mathbf{b} . Let the random matrix \mathbf{B} have rows $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$. Let $\mathcal{V} \subset \mathcal{S}^{n-1}$. If m satisfies

$$m > \frac{c'\alpha^4}{\beta^2} \ell_*(\mathcal{V})^2,$$

then with probability at least $1 - \exp(-\bar{c}\beta^2 m/\alpha^4)$, for all $\mathbf{v} \in \mathcal{V}$, we have

$$1 - \beta \leq \frac{\|\mathbf{B}\mathbf{v}\|_2^2}{m} \leq 1 + \beta,$$

where c', \bar{c} are positive absolute constants.

Recall the definitions in (31) and (33), and set $\mathcal{V} = \mathcal{W}(\mathcal{E}_k)$. Then from Lemma 5, Lemma 6 and Lemma 7, we obtain the following result: if the number of measurements

$$m > \frac{c_1 k \rho_{\max}(k)}{p^2 \beta^2 \rho_{\min}(k)} \log \frac{5en}{k}, \quad (37)$$

then with probability at least $1 - \exp(-c_2\beta^2 p^2 m/4)$, for all $\mathbf{v} \in \mathcal{E}_k$, we have

$$1 - \beta \leq \|\tilde{\mathbf{Z}}\Sigma\mathbf{v}\|_2^2 \leq 1 + \beta, \quad (38)$$

where $c_1 \triangleq 36c'c_0^4$ and $c_2 \triangleq \bar{c}/c_0^4$ are positive absolute constants.

Furthermore, by replacing \mathbf{v} with the Σ -normalized vector $\mathbf{v}/\|\Sigma\mathbf{v}\|_2$ in (38), we obtain

$$(1 - \beta)\|\Sigma\mathbf{v}\|_2^2 \leq \|\tilde{\mathbf{Z}}\Sigma\mathbf{v}\|_2^2 \leq (1 + \beta)\|\Sigma\mathbf{v}\|_2^2 \quad (39)$$

holds with probability at least $1 - \exp(-c_2\beta^2 p^2 m/4)$.

Step 2: Restricted Isometry Property. From (39) and the definitions of the k -restricted extreme eigenvalues in (15), for any k -sparse vector \mathbf{x} , we obtain that the following inequality

$$(1 - \beta)\rho_{\min}(k)\|\mathbf{x}\|_2^2 \leq \|\tilde{\mathbf{Z}}\Sigma\mathbf{x}\|_2^2 \leq (1 + \beta)\rho_{\max}(k)\|\mathbf{x}\|_2^2, \quad (40)$$

holds with probability at least $1 - \exp(-c_2mp^2\beta^2/4)$.

Recall the definitions of the parameters ξ_k , ζ_k , ϑ_k , β_k , and δ_k defined prior to Theorem 1. As in (40), the LHS and the RHS may have different deviations from one. Hence, the maximum operation and piecewise linear mappings are used in those definitions, such that after some simple substitutions and algebraic manipulations, the following inequality

$$(1 - \delta_k)\|\mathbf{x}\|_2^2 \leq \|\tilde{\mathbf{Z}}\Sigma\mathbf{x}\|_2^2 \leq (1 + \delta_k)\|\mathbf{x}\|_2^2 \quad (41)$$

holds with probability at least $1 - \exp(-c_2mp^2\beta_k^2/4)$. Collecting the results in (37) and (41), we obtain Theorem 1 for the real case.

Step 3: Generalization to the complex case. We generalize the above RIP result to the complex case. First, we show that the matrix $\tilde{\mathbf{Z}}\Sigma$ satisfies the RIP for the complex data $\mathbf{x} = \mathbf{x}^R + j\mathbf{x}^I$. With probability at least $1 - \exp(-c_2mp^2\beta_k^2/4)$, we have

$$(1 - \delta_k)\|\mathbf{x}^R\|_2^2 \leq \|\tilde{\mathbf{Z}}\Sigma\mathbf{x}^R\|_2^2 \leq (1 + \delta_k)\|\mathbf{x}^R\|_2^2,$$

$$(1 - \delta_k)\|\mathbf{x}^I\|_2^2 \leq \|\tilde{\mathbf{Z}}\Sigma\mathbf{x}^I\|_2^2 \leq (1 + \delta_k)\|\mathbf{x}^I\|_2^2.$$

Combining the above two equations yields

$$(1 - \delta_k)\|\mathbf{x}\|_2^2 \leq \|\tilde{\mathbf{Z}}\Sigma\mathbf{x}\|_2^2 \leq (1 + \delta_k)\|\mathbf{x}\|_2^2.$$

Second, we show that when the sensing matrix \mathbf{A} in our scheme is complex random matrix, it still satisfies the RIP. Let $\mathbf{A} = \mathbf{A}^R + j\mathbf{A}^I$. It is assumed that the real part \mathbf{A}^R and the imaginary part \mathbf{A}^I are independent, and have the same probability distribution. Recall that the sensing matrix $\mathbf{A} = \tilde{\mathbf{Z}}\Sigma$. For any k -sparse complex vector \mathbf{x} , we have

$$\begin{aligned}\frac{1}{2}(1 - \delta_k)\|\mathbf{x}\|_2^2 &\leq \|\mathbf{A}^R\mathbf{x}\|_2^2 \leq \frac{1}{2}(1 + \delta_k)\|\mathbf{x}\|_2^2, \\ \frac{1}{2}(1 - \delta_k)\|\mathbf{x}\|_2^2 &\leq \|\mathbf{A}^I\mathbf{x}\|_2^2 \leq \frac{1}{2}(1 + \delta_k)\|\mathbf{x}\|_2^2.\end{aligned}$$

Combining the above two equations yields the RIP in (14) for the general complex case. ■

B. Proof of Theorem 2

Proof: Clearly, we have $\rho_{\max}(1, n) = \rho_{\min}(1, n) = 1$ so the bounds are satisfied for $k = 1$. We will first prove Theorem 2 for the case $k = 2$. Subsequently, we generalize the result to arbitrary $2 \leq k < \lfloor n/2 \rfloor$. Let the two non-zero elements be $v_{s_1} = A_1 e^{j\theta_1}$ and $v_{s_2} = A_2 e^{j\theta_2}$, where $A_1^2 + A_2^2 = 1$ (because $\|\mathbf{v}\|_2 = 1$). Then from (27), and the fact that $P_{\text{ave}} = \sum_{i=1}^n \gamma_i/n$, we obtain

$$\begin{aligned}\|\mathbf{w}\|_2^2 &= \frac{1}{nP_{\text{ave}}} \sum_{i=1}^n \gamma_i + \frac{2A_1 A_2}{nP_{\text{ave}}} \sum_{i=1}^n \gamma_i \cos\left(\theta + \frac{2\pi(i-1)\Delta}{n}\right) \\ &= 1 + 2A_1 A_2 \frac{\sum_{i=1}^n a_i \gamma_i}{\sum_{i=1}^n \gamma_i},\end{aligned}\tag{42}$$

where $\theta \triangleq \theta_1 - \theta_2 \in (0, 2\pi]$, $\Delta \triangleq s_2 - s_1 \in \{1, \dots, n-1\}$, and $a_i \triangleq \cos(\theta + 2\pi(i-1)\Delta/n)$. We now set $X_i = \gamma_i$ to emphasize that the signal powers are random variables. Recall that the distributions of X_i 's are truncated Gaussian, denoted by $\mathcal{N}_{\text{tr}}(\mu, \omega)$. We consider the random variable

$$S_n \triangleq \frac{\sum_{i=1}^n a_i X_i}{\sum_{i=1}^n X_i}.\tag{43}$$

We define the Cesàro-sum of the a_i 's as

$$\bar{a}_n \triangleq \frac{1}{n} \sum_{i=1}^n a_i = \frac{1}{n} \sum_{i=1}^n \cos\left(\theta + \frac{2\pi(i-1)\Delta}{n}\right),\tag{44}$$

and note that as $n \rightarrow \infty$, the Cesàro-sum converge. Indeed, we have

$$\bar{a}_n \rightarrow \bar{a} = \frac{1}{2\pi\Delta} \int_0^{2\pi\Delta} \cos(\theta + \Delta t) dt = 0.\tag{45}$$

We now bound the probability that S_n exceeds some $t > 0$ by considering the chain of inequalities

$$\begin{aligned}
\mathbb{P}(S_n > t) &= \mathbb{P}\left(\frac{\sum_{i=1}^n a_i X_i}{\sum_{i=1}^n X_i} > t\right) \\
&\stackrel{(a)}{=} \mathbb{P}\left(\sum_{i=1}^n a_i X_i > t \sum_{i=1}^n X_i\right) \\
&\stackrel{(b)}{\leq} \mathbb{P}\left(\left\{\sum_{i=1}^n a_i X_i > t \sum_{i=1}^n X_i\right\} \cap \left\{\frac{1}{n} \sum_{i=1}^n X_i > \tau\mu\right\}\right) + \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n X_i \leq \tau\mu\right) \\
&\stackrel{(c)}{\leq} \mathbb{P}\left(\left\{\frac{1}{n} \sum_{i=1}^n a_i X_i > t\tau\mu\right\} \cap \left\{\frac{1}{n} \sum_{i=1}^n X_i > \tau\mu\right\}\right) + \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n X_i \leq \tau\mu\right) \\
&\leq \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n a_i X_i > t\tau\mu\right) + \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n X_i \leq \tau\mu\right),
\end{aligned} \tag{46}$$

where (a) is due to the fact that X_i 's are nonnegative random variables, (b) follows from the fact $\mathbb{P}(\mathcal{A}) = \mathbb{P}(\mathcal{A} \cap \mathcal{B}) + \mathbb{P}(\mathcal{A} \cap \mathcal{B}^c) \leq \mathbb{P}(\mathcal{A} \cap \mathcal{B}) + \mathbb{P}(\mathcal{B}^c)$ and (c) comes from monotonicity of measure. In the following, we bound the two terms in (46) using the theory of large deviations [17].

Define $t' \triangleq t\tau\mu$ and let s be an arbitrary non-negative number. Then from Markov's inequality, the first term in (46) can be upper bounded as follows

$$\mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n a_i X_i > t'\right) \leq \exp(-nst') \mathbb{E}\left[\exp\left(\sum_{i=1}^n sa_i X_i\right)\right], \tag{47}$$

which implies by the independence of the X_i 's that

$$\frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n a_i X_i > t'\right) \leq -st' + \frac{1}{n} \sum_{i=1}^n \log \mathbb{E}[\exp(sa_i X_i)]. \tag{48}$$

To bound the sum in (48), we find the cumulant-generating function (CGF) of $X \sim \mathcal{N}_{\text{tr}}(\mu, \omega^2)$ in terms of a Gaussian with mean μ and variance ω^2 . By simple algebraic manipulations, we have

$$\log \mathbb{E}[\exp(sX)] = \mu s + \frac{1}{2} \omega^2 s^2 + \varphi(\mu, \omega, s), \tag{49}$$

where $\varphi(\mu, \omega, s) \triangleq \log(1 - Q(\mu/\omega + \omega s)) - \log(1 - Q(\mu/\omega))$. We note that given that (μ, ω) is a positive pair of numbers, $s \mapsto \varphi(\mu, \omega, s)$ for $s \geq 0$ is concave, because $s \mapsto -Q(\mu/\omega + \omega s)$ (for $\mu/\omega > 0$) and $t \mapsto \log(1+t)$ are both concave and the latter function is non-decreasing. Moreover, $s \mapsto \varphi(\mu, \omega, s)$ is continuous for each positive (μ, ω) pair, because every concave function on an open set is continuous. Note that $\varphi(\mu, \omega, 0) = 0$.

Substituting the CGF of the truncated Gaussian distribution in (49) into (48) yields

$$\begin{aligned}
& \frac{1}{n} \log \mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n a_i X_i > t' \right) \\
& \leq -st' + \mu s \bar{a}_n + \frac{\omega^2 s^2}{2n} \sum_{i=1}^n a_i^2 + \frac{1}{n} \sum_{i=1}^n \varphi(\mu, \omega, a_i s) \\
& \stackrel{(a)}{=} -st' + \mu s \bar{a}_n + \frac{\omega^2 s^2}{4} + \frac{\omega^2 s^2}{4n} \sum_{i=1}^n \cos \left(2\theta + \frac{4\pi\Delta(i-1)}{n} \right) + \frac{1}{n} \sum_{i=1}^n \varphi(\mu, \omega, a_i s) \\
& \stackrel{(b)}{\leq} -st' + \mu s \bar{a}_n + \frac{\omega^2 s^2}{4} + \frac{\omega^2 s^2}{4n} \sum_{i=1}^n \cos \left(2\theta + \frac{4\pi\Delta(i-1)}{n} \right) + \varphi \left(\mu, \omega, \frac{s}{n} \sum_{i=1}^n a_i \right), \tag{50}
\end{aligned}$$

where (a) comes from the definition of a_i and the double-angle formula for the cosine, and (b) follows the fact $\varphi(\mu, \omega, s)$ is concave in s for any positive (μ, ω) pair.

Taking the limsup on both sides of (50) and using the definition of \bar{a}_n yields

$$\begin{aligned}
& \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n a_i X_i > t' \right) \\
& \stackrel{(a)}{\leq} -st' + \frac{\omega^2 s^2}{4} + \frac{\omega^2 s^2}{16\pi\Delta} \int_0^{4\pi\Delta} \cos(2\theta + t) dt + \limsup_{n \rightarrow \infty} \varphi(\mu, \omega, \bar{a}_n s) \\
& \stackrel{(b)}{=} -st' + \frac{\omega^2 s^2}{4} + \limsup_{n \rightarrow \infty} \varphi(\mu, \omega, \bar{a}_n s) \\
& \stackrel{(c)}{=} -st' + \frac{\omega^2 s^2}{4} \triangleq f(s), \tag{51}
\end{aligned}$$

where (a) follows from Riemann sums, (b) comes from the fact cosine has zero mean over an integer number of periods (note $\Delta \in \mathbb{Z}$) and (c) follows from the continuity of $\varphi(\mu, \omega, s)$ and (45). Note that the minimum $f(s)$ in (51) is $f(s^*) = -\tau^2 d^2 t^2$ (attained at $s^* = 2t'/\omega^2$). Hence,

$$\mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n a_i X_i > t' \right) \dot{\leq} \exp \left[-n\tau^2 d^2 t^2 \right]. \tag{52}$$

The second term in (46) can be bounded using standard techniques from the large deviations theory [17] (Cramér's theorem) and along the same lines as the derivation above. As such we have

$$\begin{aligned}
\mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n X_i \leq \tau \mu \right) & \stackrel{(a)}{\leq} \exp \left[-n \left(s\mu(1-\tau) - \omega^2 s^2/2 - \varphi(\mu, \omega, -s) \right) \right] \\
& \stackrel{(b)}{\leq} \exp \left[-n \left(s\mu(1-\tau) - \omega^2 s^2/2 \right) \right],
\end{aligned}$$

where (a) follows from using the CGF of X_i in (49), and (b) follows from the fact that $\varphi(\mu, \omega, -s) \leq 0$ for all $s \geq 0$. Hence, setting $s \triangleq \mu(1 - \tau)/\omega^2$, we have

$$\mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n X_i \leq \tau \mu \right) \leq \exp \left[-n(1 - \tau)^2 d^2 / 2 \right]. \quad (53)$$

Combining the two terms in (46), we have from (52) and (53) and the largest-exponent-dominates principle that

$$\mathbb{P}(S_n > t) \leq \exp \left[-n \min \left\{ \tau^2 d^2 t^2, (1 - \tau)^2 d^2 / 2 \right\} \right] \quad (54)$$

Since $\tau > 0$ is a free parameter, we can set it to be $\tau^* \triangleq \frac{1}{1 + \sqrt{2t}}$. Substituting τ^* into (54) yields

$$\mathbb{P}(S_n > t) \leq -nd^2 \tilde{t}^2. \quad (55)$$

where $\tilde{t} \triangleq t/(1 + \sqrt{2t})$. By symmetry, we can also conclude that

$$\mathbb{P}(S_n < -t) \leq -nd^2 \tilde{t}^2. \quad (56)$$

Recall that $\rho_{\max}(k, n)$ is the maximum value of $\|\mathbf{w}\|_2^2 = \|\Sigma \mathbf{v}\|_2^2$ over all unit-norm k -sparse vectors \mathbf{v} . From (42), $\|\mathbf{w}\|_2^2$ depends only on $A_1 A_2$. Note that $0 < A_1 A_2 \leq 1/2$ because $\sqrt{A_1 A_2} \leq (A_1 + A_2)/2$. We set $A_1 A_2 = 1/2$, whence $\|\mathbf{w}\|_2$ attains its maximum value. From (42),

$$\begin{aligned} \mathbb{P}(\rho_{\max}(2, n) > 1 + t) &\leq \exp \left[-nd^2 \tilde{t}^2 \right], \\ \mathbb{P}(\rho_{\min}(2, n) < 1 - t) &\leq \exp \left[-nd^2 \tilde{t}^2 \right]. \end{aligned} \quad (57)$$

Having proved the result for the $k = 2$ case, we now generalize it to the case where $k > 2$. Set the non-zero elements of the vector \mathbf{v} to be $v_{s_q} = A_q e^{j\theta_q}$, $q = 1, \dots, k$, where $\sum_{q=1}^k A_q^2 = 1$. Equation (27) can be written as

$$\begin{aligned} \|\mathbf{w}\|_2^2 &= \frac{1}{nP_{\text{ave}}} \sum_{i=1}^n \gamma_i \left(1 + \sum_{q=1}^k \sum_{l=1, l \neq q}^k A_q A_l \cos \left(\theta_{q,l} + \frac{j2\pi(i-1)\Delta_{q,l}}{n} \right) \right) \\ &= 1 + \sum_{q=1}^k \sum_{l=1, l \neq q}^k A_q A_l \sum_{i=1}^k \cos \left(\theta_{q,l} + \frac{2\pi(i-1)\Delta_{q,l}}{n} \right), \\ &= 1 + \sum_{q=1}^k \sum_{l=1, l \neq q}^k A_q A_l S_n^{q,l} \triangleq 1 + B_n, \end{aligned} \quad (58)$$

where $S_n^{q,l}$ is defined as in (43) but involving the q -th and the l -th nonzero elements of \mathbf{v} , i.e.,

$\theta_{q,l} = \theta_q - \theta_l$, and $\Delta_{q,l} = s_l - s_q$. On the other hand, we can bound B_n^2 as follows

$$\begin{aligned}
B_n^2 &\stackrel{(a)}{\leq} \left(\sum_{q=1}^k \sum_{l=1, l \neq q}^k A_q^2 A_l^2 \right) \left(\sum_{q=1}^k \sum_{l=1, l \neq q}^k (S_n^{q,l})^2 \right) \\
&= \left(\left(\sum_{q=1}^k A_q^2 \right)^2 - \sum_{q=1}^k A_q^4 \right) \left(\sum_{q=1}^k \sum_{l=1, l \neq q}^k (S_n^{q,l})^2 \right) \\
&\stackrel{(b)}{\leq} \left(1 - \frac{1}{k} \left(\sum_{q=1}^k A_q^2 \right)^2 \right) \left(\sum_{q=1}^k \sum_{l=1, l \neq q}^k (S_n^{q,l})^2 \right) \\
&= \frac{k-1}{k} \sum_{q=1}^k \sum_{l=1, l \neq q}^k (S_n^{q,l})^2,
\end{aligned} \tag{59}$$

where (a) comes from the Cauchy-Schwartz inequality and (b) comes from the basic inequality relating the arithmetic and quadratic means, namely $1/M \sum_{j=1}^M \alpha_j \leq (1/M \sum_{j=1}^M \alpha_j^2)^{1/2}$.

Now, given any $t > 0$, we can bound the probability that $|B_n|$ exceeds t as follows:

$$\begin{aligned}
\mathbb{P}(|B_n| > t) &= \mathbb{P}(B_n^2 > t^2) \\
&\stackrel{(a)}{\leq} \mathbb{P}\left(\sum_{q=1}^k \sum_{l=1, l \neq q}^k (S_n^{q,l})^2 > \frac{kt^2}{k-1}\right) \\
&\leq \mathbb{P}\left(\max_{l \neq q} (S_n^{q,l})^2 > \frac{t^2}{(k-1)^2}\right) \\
&\stackrel{(b)}{\leq} \sum_{q=1}^k \sum_{l=1, l \neq q}^k \mathbb{P}\left((S_n^{q,l})^2 > \frac{t^2}{(k-1)^2}\right) \\
&= \sum_{q=1}^k \sum_{l=1, l \neq q}^k \mathbb{P}\left(S_n^{q,l} > \frac{t}{k-1}\right),
\end{aligned} \tag{60}$$

where (a) comes from (59) and monotonicity of measure and (b) comes from the union bound. Applying the result for $k = 2$ in (57) to (60), we have

$$\mathbb{P}(|B_n| > t) \leq k(k-1) \exp[-nd^2 E(k, t)^2], \tag{61}$$

where the exponent is $E(k, t) \triangleq t/(k-1 + \sqrt{2}t)$. Recall the definition of $\rho_{\max}(k, n)$ in (15). From (58) and (61), we conclude that

$$\mathbb{P}(\rho_{\max}(k, n) > 1 + t) \leq \exp[-nd^2 E(k, t)^2]. \tag{62}$$

The analysis of $\mathbb{P}(\rho_{\min}(k, n) < 1 - t)$ proceeds *mutatis mutandis*. This completes the proof. ■

REFERENCES

- [1] A. Kansal, J. Hsu, S. Zahedi, and M. B. Srivastava, "Power management in energy harvesting sensor networks," *ACM Trans. Embed. Comput. Syst.*, vol. 6, Sept. 2007.
- [2] C. K. Ho and R. Zhang, "Optimal energy allocation for wireless communications with energy harvesting constraints," *IEEE Trans. Signal Process.*, vol. 60, pp. 4808–4818, Sept. 2012.
- [3] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, pp. 21–30, Mar. 2008.
- [4] J. D. Haupt and R. D. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inf. Theory*, vol. 52, pp. 4036–4048, Sept. 2006.
- [5] S. Aeron, M. Zhao, and V. Saligrama, "Information theoretic bounds to sensing capacity of sensor networks under fixed SNR," in *IEEE Inf. Th. Workshop*, (Lake Tahoe, CA, USA), pp. 84–89, Sept. 2007.
- [6] R. Rana, W. Hu, and C. T. Chou, "Energy-aware sparse approximation technique (EAST) for rechargeable wireless sensor networks," in *European Conf. on Wireless Sensor Networks*, (Coimbra, Portugal), pp. 306–321, Feb. 2010.
- [7] T. Xue, X. Dong, and Y. Shi, "A multiple access scheme based on multi-dimensional compressed sensing," in *IEEE Int. Conf. on Commun. (ICC)*, (Ottawa, Canada), pp. 3832–3836, Jun. 2012.
- [8] F. Fazel, M. Fazel, and M. Stojanovic, "Random access compressed sensing for energy-efficient underwater sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 29, pp. 1660–1670, Sept. 2011.
- [9] W. Bajwa, *New information processing theory and methods for exploiting sparsity in wireless systems*. PhD thesis, University of Wisconsin-Madison, 2009.
- [10] A. M. Tulino and S. Verdú, *Random matrix theory and wireless communications*. Hanover, MA, USA: now Publishers Inc. Press, 2004.
- [11] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4203–4215, Dec. 2005.
- [12] R. Baraniuk, M. Davenport, R. D. Vore, and M. Wakin, "A simple proof of the restricted isometry property," *Constr. Approx.*, vol. 28, no. 3, pp. 253–263, 2008.
- [13] Y. C. Eldar and G. Kutyniok, *Compressed sensing: Theory and applications*. Cambridge Univ. Press, 2012.
- [14] S. Mendelson, A. Pajor, and N. T. Jaegermann, "Uniform uncertainty principle for Bernoulli and sub-Gaussian ensembles," *Constr. Approx.*, vol. 28, pp. 277–289, 2008.
- [15] T. T. Cai, M. Wang, and G. Xu, "New bounds for restricted isometry constants," *IEEE Trans. Inf. Theory*, vol. 56, pp. 4388–4394, Sept. 2010.
- [16] M. A. Davenport, "The pros and cons of compressive sensing for wideband signal acquisition: Noise folding versus dynamic range," *IEEE Trans. Signal Process.*, vol. 60, pp. 4628–4642, Sept. 2012.
- [17] A. S. Dembo and O. Zeitouni, *Large deviation techniques and applications*. Springer Press, 1998.
- [18] E. V. D. Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *Proc. of Soc. Ind. Appl. Math.*, vol. 31, no. 2, pp. 890–912, 2008.
- [19] S. Zhou, "Restricted eigenvalue conditions on sub-Gaussian random matrices." Website, 2009. <http://arxiv.org/abs/0912.4045v2>.